Recognition of Manipulation Sequences by Human Hand Based on Support Vector Machine

Kazuya Matsuo Kouji Murakami Tsutomu Hasegawa Ryo Kurazume
Graduate School of Information Science and Electrical Engineering,
Kyushu University
744 Motooka, Nishi-Ku, Fukuoka 819-0395 JAPAN
{matsuo, mkouji, hasegawa, kurazume}@irvs.is.kyushu-u.ac.jp

***Abstract*— This paper describes a method of recognizing a manual task executed by a human hand by using the Support Vector Machine (SVM). We define several task states which are segmented from the continuous motion of human fingers in the context of an object manipulation. Based on margins of SVMs, the method constructs a binary decision tree which most effectively classifies and symbolizes the task state from joint angle trajectories of human fingers as input. The binary decision tree constructed by our method has been evaluated through experiments of recognizing the task states during a valve manipulation.**

***Index Terms*— multi-jointed multi-fingered robotic hand, support vector machine**

I. Introduction

A multi-jointed multi-fingered robotic hand is a potentially dexterous hand like a human hand [1]. However, it is very difficult for us to manually and directly write down a motion program of multi-fingers moving synchronously and cooperatively to execute a task. Teaching-by-showing is an alternative approach: a motion of human fingers is measured and recorded, then it is somehow transformed into a motion program of a robotic hand. This paper proposes a method of continuous motion recognition of human fingers executing a manipulation task. The method uses the Support Vector Machine (SVM).

There are usually many differences of structure between a human hand and a robotic hand: number of fingers, number of joints, length of a finger link and so on. Therefore it is not straightforward to utilize the data obtained in Teaching-by-Showing by a human worker. Direct mapping of joint angle trajectories of the human hand onto the robotic hand will not execute the same manual task. Another method is to generate joint angle trajectories of the robotic hand from the fingertip trajectories of the human hand in the Cartesian coordinate system by solving inverse kinematics of the robotic hand. This may easily fail in keeping stable grasp of an object because possible direction and magnitude of the operating force differs due to the difference of the finger configuration even if the fingertip position is the same. Moreover, some joints may go over the range of rotation. The force control is essential to overcome error in human motion measurement. Wang et al. propose a method of modifying fingertip positions of the robot according to the limit of rotation when the fingertip trajectories of the human hand is transformed and processed [2].

A task program for the robotic hand would be autonomously generated, if the actual task being executed by a human is recognized from the motion data of a human hand obtained in Teaching-by-Showing process. Related work on human motion recognition has been reported in [3][4][5]: the continuous human motion is segmented, recognized, and symbolized according to the meaning of the particular motion in the context of the task. A manual task by the human hand is represented by a sequence of symbols each representing particular manipulation. Adequately abstracted symbols enable to develop corresponding motion primitives feasible by a robot hand. Then the robot system would autonomously be able to perform various tasks by executing a sequence of corresponding motion primitives when a sequence of symbolic descriptions of a task performed by a human is given.

Several methods have been proposed to recognize a manipulation task performed by a human worker: they are classification based on Euclid distance between motion data [6], Hidden Markov Model (HMM) [7][8], and Dynamic Programming matching (DP matching) [9]. The classification by Euclid distance lacks generality to be applicable to an object of different size or to a different subject performing the same task. HMM and DP matching achieve a high recognition rate by using state transition model of manual tasks represented by finite state automata. However, it would be difficult to implement so many state transition models corresponding to various manipulation works in our daily human life. Each time a new task is implemented, the state transition model must be redesigned based on transition probability of the new task from previously implemented tasks. In contrast to these, instantaneous recognition method recognizes the state of a manipulation task from input motion data without relying on the past results of recognition of time series of motion data. Therefore, addition of new tasks to be recognized would be easier. Instantaneous recognition will not be influenced by past error of recognition, although the recognition itself would face with difficulty because the context information of a time series of tasks is not available.

This paper proposes an instantaneous recognition method

of a manipulation task for a robotic hand. Based on the SVM, the method recognizes a manual task executed by a human hand from the joint angle trajectories of human fingers as input, and symbolizes the task. We define several task states by segmenting the continuous motion of human fingers within the context of manipulation of an object. The task states and a hand task are related as followings:

$$U = \{S_i \mid 1 \leq i \leq N,\ N : the\ number\ of\ task\ states\} \tag{1}$$

$$V = U^+ \tag{2}$$

where S_i is task state, U is the universal set of task states, V is set of hand tasks, and U^+ is the set of all strings over elements in U except the empty string.

This paper is structured as follows. Section II describes the input device and input data for recognition of human fingers executing a manipulation task. A method of recognition is proposed using a binary decision tree which consists of SVMs in Section III. Section IV shows the performed experiments and analyzes the obtained results. Section V is a conclusion of this paper.

II. Measurement System

A. Input Device

We use Cyber Glove (CG1802-R) produced by Immersion Corporation as the input device for measuring the joint angles of a human hand. Its appearance and specifications are shown in Fig.1 and TABLE I respectively. Cyber Glove has capability to measure eighteen joints of a human hand. In order to recognize a task state, we use the angles of sixteen joints except for two joints of a wrist. The positions of sixteen joints are shown in Fig.2. The thumb's proximal joint has two DOFs. Each of every other joint has one DOF.

Fig. 1. Cyber Glove.

TABLE I
Specifications of Cyber Glove (CG1802-R).

Number of Sensors	18
Sensor Resolution	0.5 degrees
Interface	RS-232
Maximum Data Rate	115.2 kbaud

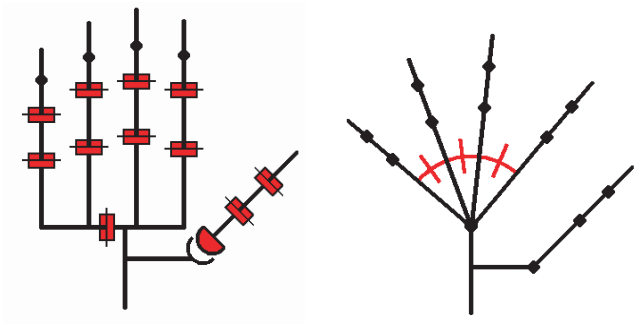

Fig. 2. Position of measured joints of right hand.

B. Input Data

We use a joint angle and an angular velocity of each of sixteen joints for recognizing a task state. Input is 32 dimensional vector. The angular velocity is derived as time-difference of the joint angle. The angular velocities are used for recognizing a task state with a dynamic finger motion.

III. Grasp Classification Based on the Support Vector Machines

A. Support Vector Machine

We recognize task state based on the SVM. The SVM is a supervised machine learning method used for classification. Training data group of the SVM consists of positive data group and negative data group. The SVM constructs a separating hyperplane with positive data group and negative data group (Fig.3). The separating hyperplane maximises the distance between the hyperplane and each data group. The distance is called 'margin'. Therefore, the SVM is known as a maximum margin classifier. The key features of the SVM are high generalization ability, the absence of local optimum and high performance of recognition with rather small number of training data.

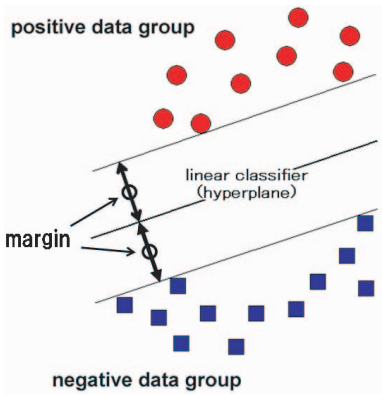

Fig. 3. Maximum-margin hyperplane for a SVM trained with samples from two classes.

B. Binary Decision Tree

Our method constructs a binary decision tree which consists of SVMs. In the binary decision tree, input data are joint angles and angular velocities of joints of human fingers. Output is a corresponding state of a task.

Classification at each decision node is implemented by an SVM. Each terminal node is labeled by a task state. In this method, a recognition error does not affect a subsequent recognition result, because the method does not use a previous recognition result.

The binary decision tree is derived from repeating division of a set of task states. U is the universal set of task states (Eq.(1)). The decision node which divides U into U_1 and U_2 is constructed (Fig.4). Both U_1 and U_2 are subsets of U (Eq.(3)). U_1 and U_2 are recursively divided into subsets respectively. When all subsets consist of one task state, the binary decision tree is completed.

$$U_1 \subset U, \quad U_2 = U - U_1 \tag{3}$$

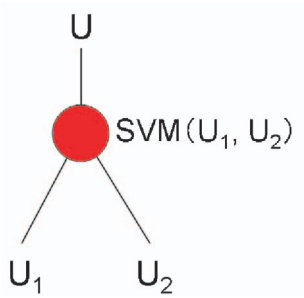

Fig. 4. A constructed decision node for a set of task states (U).

The decision node is shown as a circle in Fig.4. $SVM(U_1,\ U_2)$ means a Support Vector Machine whose positive data group and negative data group are training data group of U_1 and U_2 respectively. Training data group of U_i is the union of training data groups of task states which are included in U_i $(i = 1, 2)$. Training data group of each task state is collected through the training demonstration of a manipulation task by a subject.

The structure of a binary decision tree depends on dividing process. When task states are to be recognized, possible structure of the binary decision tree is not unique. For example, when five task states are to be recognized, the number of possible structure of the binary decision trees is one hundred five. Through preliminary experiments, we found out that structure of the binary decision tree affects recognition rate of the task state. So, we propose a method of constructing a binary decision tree with high recognition rate by using margins of SVMs.

C. Constructing a binary decision tree based on margins of SVMs

Classification at an upper decision node of the tree affects classifications at lower decision nodes. Therefore, a binary decision tree having high classification ability should have those SVMs having high classification ability in its upper nodes. The SVM has high classification ability if it has high generalization ability. The larger the margin of an SVM is, the higher the generalization ability is. The algorithm we propose here constructs the binary decision tree according to the margins of SVMs. The algorithm is explained as follows.

(1) $|U| \rightarrow N$. $U = \{S_i \mid 1 \leq i \leq N,\ N : the\ number\ of\ task\ states\}$.
(2) if $N = 1$ then end.
(3) U_1 is a proper subset of U $(U_1 \subset U)$. U_2 is the absolute complement of U_1 $(U_2 = U - U_1)$. Select U_1max and U_2max from possible proper subsets so that the margin of $SVM(U_1max, U_2max)$ be maximized.
(4) Construct a binary decision tree. The decision node of the tree is labeled by $SVM(U_1max, U_2max)$. Two terminal nodes of the tree are labeled by U_1max and U_2max respectively.
(5) $U_1max \rightarrow U$, $U_2max \rightarrow U$.
go to (1).

Margin of upper SVM is larger than that of lower SVM in the tree derived from this method.

IV. Experiment

We have observed several typical tasks by a human hand, and defined five task states which are segmented from the continuous hand motion. They are shown in Fig.5.

1) S_{open} : open hand without touching an object.
2) S_{pull} : pulling an object.
3) S_{grasp} : grasping an object.
4) $S_{rotate-right}$: rotating an object clockwise.
5) $S_{rotate-left}$: rotating an object counterclockwise.

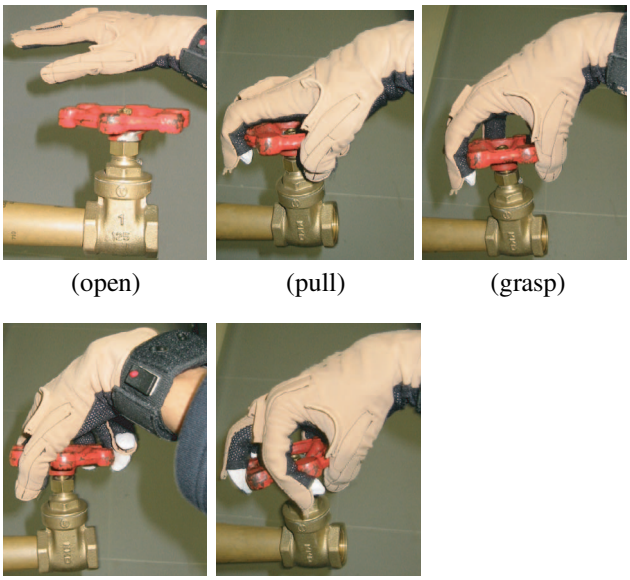

Fig. 5. Appearances of task states.

During a rotation task, a human has to regrasp the object due to rotational limit of joint angle. Thus the rotation task consists of rotating and regrasping. Rotating and regrasping are repeated alternately.

The valve manipulation is one of typical tasks expressed as a sequence of task states defined above. We have constructed a binary decision tree to recognize a task state

during the valve manipulation. The diameter of the valve is $78mm$. The valve is fixed to a work table.

Positive data group of each task state is collected through the training demonstration by a subject (subject-A). Binary decision tree is derived from these positive data groups. The generated tree is shown in Fig.6. Decision nodes are shown as the circles. Training data group of each SVM consists of several positive data groups. The numbers of the positive data in S_{open}, S_{pull}, S_{grasp}, $S_{rotate-right}$ and $S_{rotate-left}$ are 38, 32, 39, 57 and 64 respectively. The SVM nodes are allocated according to the margin. At first, S_{pull} is separated from the other task states. Note that the state allocation in the binary decision tree does not necessarily coincide with our intuition. For example, the nearest task state of $S_{rotate-left}$ is not $S_{rotate-right}$ but S_{grasp}.

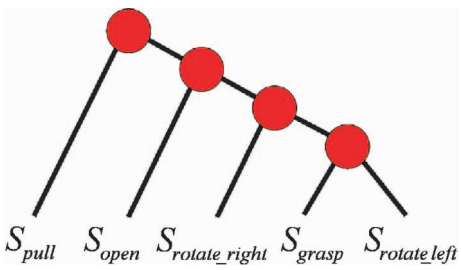

Fig. 6. Binary decision tree derived by the proposed method.

We then evaluate the generated decision tree described above. The same subject performed the experiment of the valve manipulation composed of a sequence of five task states, and the evaluation data is obtained.

The recognition of the task states can be accomplished in real time. Sampling rate of Cyber Glove is 21[Hz]. The result of recognition on these evaluation data is shown in Fig.7. Vertical axis shows task state. Horizontal axis shows the elapsed time. The true value is shown as the dotted line, while the recognition result is shown as the real line. The recognition rate of each task state is shown in Fig.8. Recognition rate is obtained as the percentage of the number of successful data to the number of input data. In the experiment, the number of input data is one thousand seventy-seven, and the recognition rate is 95.3%.

The most frequent error is to recognize $S_{rotate-left}$ as S_{grasp}. The reason that the misclassification occurs is explained as the following. When a human regrasps the valve during the rotation task, he stops his hand momentarily. The recognition result is S_{grasp} when a human finger stops its motion. Because subject-A is right-handed, the misclassification occurs frequently especially during the task of $S_{rotate-left}$. In addition, the misclassification occurs on the time a transition between task states occurs.

ideal \ result	S_{open}	S_{pull}	S_{grasp}	S_{rotate_right}	S_{rotate_left}	Recognition rate [%]
S_{open}	139	0	1	3	13	89.1
S_{pull}	0	100	0	0	0	100.0
S_{grasp}	0	3	461	0	3	98.7
S_{rotate_right}	0	0	2	171	1	98.3
S_{rotate_left}	0	0	25	0	155	86.1

Fig. 8. Recognition rate of the proposed method.

We thoroughly investigated recognition rate of all the possible one hundred and five trees against five task states defined above. We use the same training data and the evaluation data with previous experiment. Histogram of recognition rate is shown in Fig.9. Vertical axis shows frequency. Horizontal axis shows recognition rate. The worst recognition rate is 71.2%. The best recognition rate is 95.3%. The tree derived by our method has the best recognition rate.

For comparison, we investigated recognition rates of three conventional methods: neural network (NN), C4.5 algorithm (C4.5), and k-nearest neighbor algorithm (k-NN). These methods are implemented by Weka[10] which is a collection of machine learning algorithms for data mining

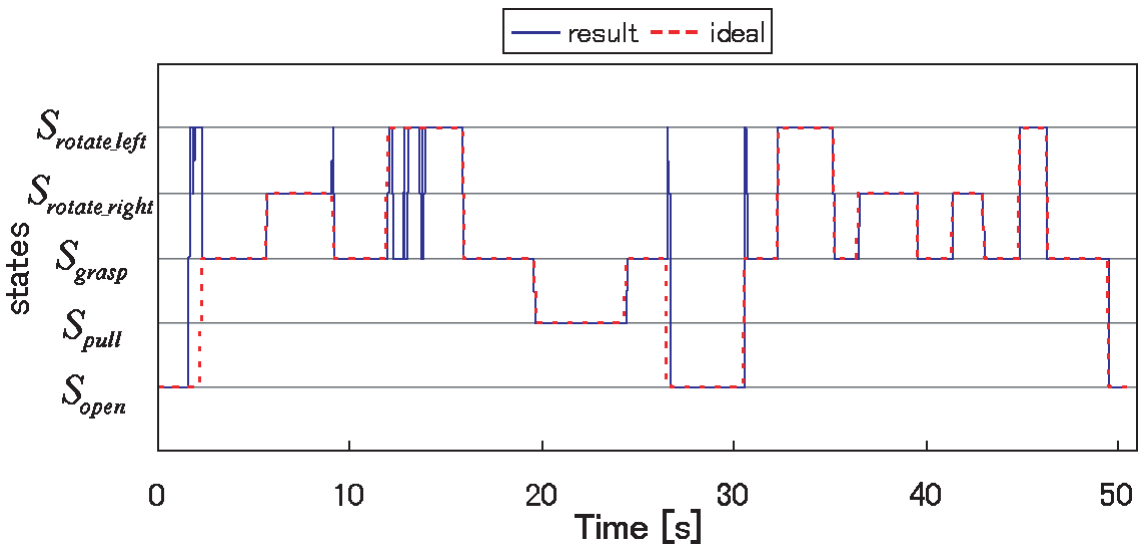

Fig. 7. Recognition result of the proposed method.

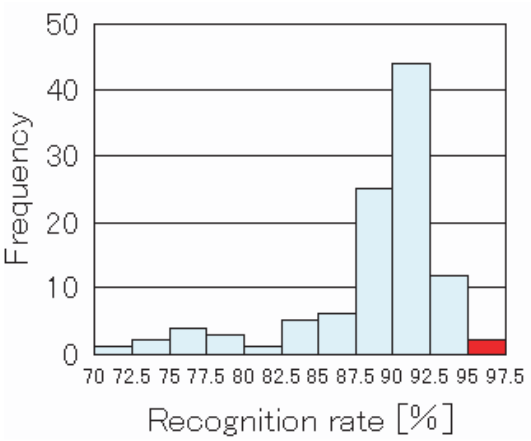

Fig. 9. Histogram on recognition rate of all possible trees.

tasks. Weka's default values are used as parameters of each method. For example, the number 'k' of neighbors to use in k-NN is 1. Parameters of the neural network are shown in TABLE II.

TABLE II

PARAMETERS OF NEURAL NETWORK

Momentum α	0.2
Learning Rate η	0.3
The number of hidden layers	1
The number of hidden units	The sum of input units and output units

The result of the comparison experiment is shown in TABLE III. Recognition rates of NN and k-NN are 95.5% and 94.6% respectively. Joint angle trajectories of the human fingers will not change significantly if the object to be manipulated is the same. Therefore, recognition rates of NN and k-NN are also high. However, when size of a manipulated object is changed, both methods may fail in recognition easily due to the large changes of joint angles of a human hand.

TABLE III

RECOGNITION RATE (THE PROPOSED METHOD AND THREE CONVENTIONAL METHODS, DIAMETER OF VALVE IS $78mm$).

task state	the proposed method	NN	C4.5	k-NN
all	95.3% (1026/1077)	95.5%	73.0%	94.6%
open	89.1% (139/156)	93.6%	96.2%	97.4%
pull	100.0% (100/100)	100.0%	100.0%	100.0%
grasp	98.7% (461/467)	98.7%	56.7%	91.6%
right	98.3% (171/174)	98.3%	71.8%	94.3%
left	86.1% (155/180)	83.3%	81.1%	97.2%

NN: neural network
C4.5: C4.5 algorithm
k-NN: k-nearest neighbor algorithm

We used two different cylinders to investigate generalization ability against change of diameter of the valve. The diameter of each cylinder is $50mm$ and $122mm$ respectively. The evaluation data group for each cylinder is obtained through the demonstration by the subject-A. The training data group of each task state is obtained through the demonstration with the valve whose diameter is $78mm$. The results of recognition are shown in TABLEs IV and V.

TABLE IV

RECOGNITION RATE (DIAMETER OF CYLINDER IS $50mm$).

task state	the proposed method	NN	C4.5	k-NN
all	93.8% (534/569)	96.0%	47.5%	90.5%
open	97.4% (75/77)	100.0%	94.8%	96.1%
pull	100.0% (90/90)	100.0%	100.0%	100.0%
grasp	95.2% (140/147)	95.2%	0.7%	93.2%
right	91.6% (109/119)	94.1%	68.9%	94.1%
left	88.2% (120/136)	93.4%	17.6%	75.0%

TABLE V

RECOGNITION RATE (DIAMETER OF CYLINDER IS $122mm$).

task state	the proposed method	NN	C4.5	k-NN
all	90.8% (580/639)	80.3%	60.9%	73.6%
open	96.0% (97/101)	100.0%	100.0%	100.0%
pull	96.2% (127/132)	92.4%	97.0%	72.0%
grasp	89.5% (145/162)	72.2%	40.1%	93.8%
right	75.2% (82/109)	44.0%	38.5%	34.9%
left	95.6% (129/135)	92.6%	39.3%	62.2%

When the diameter of the cylinder is $50mm$, recognition rates of the proposed method, NN and k-NN are 93.8%, 96.0% and 90.5% respectively. When the diameter of the cylinder is $122mm$, recognition rates of the proposed method, NN and k-NN are 90.8%, 80.3% and 73.6% respectively. Recognition rates of NN and k-NN decrease more than 15% and 20% at the maximum when diameter of the cylinder is changed. On the other hand, the decrease of recognition rate of our method is less than 5%.

Another person (subject-B) manipulated the valve and the cylinders to acquire the evaluation data groups. Generalization ability against change of size of a human hand is investigated by using these data groups. The training data groups are the same as those of previous experiments. The results of recognition are shown TABLEs VI to VIII.

TABLE VI

RECOGNITION RATE (DIAMETER OF VALVE IS $78mm$, SUBJECT B).

task state	the proposed method	NN	C4.5	k-NN
all	86.4% (678/785)	85.5%	79.0%	85.4%
open	96.9% (123/127)	100.0%	97.6%	100.0%
pull	100.0% (107/107)	98.1%	99.1%	97.2%
grasp	96.2% (350/364)	97.5%	84.3%	87.1%
right	34.9% (22/63)	30.2%	0.0%	38.1%
left	86.1% (76/124)	52.4%	66.9%	79.0%

TABLE VII

RECOGNITION RATE (DIAMETER OF CYLINDER IS $50mm$, SUBJECT B).

task state	the proposed method	NN	C4.5	k-NN
all	77.4% (503/650)	68.8%	44.3%	77.5%
open	90.2% (74/82)	93.9%	85.4%	81.7%
pull	100.0% (64/64)	100.0%	100.0%	95.3%
grasp	94.3% (263/279)	64.5%	21.9%	97.8%
right	10.5% (12/114)	35.1%	0.9%	19.3%
left	81.1% (90/111)	77.5%	82.9%	73.0%

TABLE VIII

RECOGNITION RATE (DIAMETER OF CYLINDER IS $122mm$, SUBJECT B).

task state	the proposed method	NN	C4.5	k-NN
all	79.2% (624/788)	76.8%	55.7%	69.8%
open	92.1% (70/76)	96.1%	96.1%	100.0%
pull	100.0% (134/134)	100.0%	100.0%	97.8%
grasp	94.2% (306/325)	98.8%	65.5%	94.2%
right	0.0% (0/129)	0.0%	0.0%	0.0%
left	91.9% (114/124)	62.1%	15.3%	29.8%

When the manipulating parameters are not changed through the training process and the evaluation process, the recognition rates of the proposed method, NN and k-NN are more than 90%. The manipulating parameters indicate size of a subject's hand and diameter of a manipulated valve. However, recognition rates of NN and k-NN decrease about 25% at the maximum when manipulating parameters are changed. On the other hand, the decrease of recognition rate of our method is less than 18% when manipulating parameters are changed.

V. CONCLUSION

We have proposed a method of recognizing a manual task executed by a human hand based on a binary decision tree which consists of SVMs. The binary decision tree constructed by our method recognizes states of the task from joint angle trajectories of human fingers as input. Through experiments of recognizing task states in a manipulation of a valve by a human hand, the decision tree showed higher generalization ability against change of diameter of a valve and size of a subject's hand than that of three conventional methods: neural network, C4.5 algorithm, and k-nearest neighbor algorithm.

REFERENCES

[1] H. Kawasaki, H. Shimomura and Y. Shimizu: "Educational-industrial complex development of an anthropomorphic robot hand 'Gifu hand' ", Advanced Robotics, Vol. 15, No. 3, pp. 357-363 (2001)

[2] H. Wang, K. H. Low, M. Y. Wang and F. Gong: "A Mapping Method for Telemanipulation of the Non-Anthropomorphic Robotic Hands with Initial Experimental Validation", Proceedings of the 2005 IEEE International Conference on Robotics and Automation, pp. 4229-4234 (2005)

[3] S. B. Kang and K. Ikeuchi: "Toward Automatic Robot Instruction from Perception - Mapping Human Grasps to Manipulator Grasps", IEEE Transactions on Robotics and Automation, Vol. 13, No. 1, pp. 81-95 (1997)

[4] K. Ogawara, J. Takamatsu, H. Kimura and K. Ikeuchi: "Generation of a task model by integrating multiple observations of human demonstrations", Proceedings of the 2002 IEEE International Conference on Robotics and Automation, pp. 1545-1550 (2002)

[5] T. Mori, Y. Segawa, M. Shimosaka and T. Sato: "Hierarchical Recognition of Daily Human Actions Based on Continuous Hidden Markov Models", Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 779-784 (2004)

[6] J. Aleotti and S. Caselli: "Grasp Recognition in Virtual Reality for Robot Pregrasp Planning by Demonstration", Proceedings of the 2006 IEEE International Conference on Robotics and Automation, pp. 2801-2806 (2006)

[7] K. Bernardin, K. Ogawara, K. Ikeuchi and R. Dillmann: "A Sensor Fusion Approach for Recognizing Continuous Human Grasping Sequences Using Hidden Markov Models", IEEE Transactions on Robotics, Vol. 21, No. 1, pp. 47-57 (2005)

[8] S. Ekvall and D. Kragic: "Grasp Recognition for Programming by Demonstration", Proceedings of the 2005 IEEE International Conference on Robotics and Automation, pp. 760-765 (2005)

[9] M. Kondo, J. Ueda, Y. Matsumoto and T. Ogasawara: "Perception of Human Manipulation Based on Contact State Transition", Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots, TA1-D3 (2004)

[10] Ian H. Witten, Eibe Frank: "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations", Morgan Kaufmann Publishers (1999)